Biostatistics I – Introduction to Statistics and Experimental Design

Grace E. Kissling, Ph.D.
Biostatistics Branch

National Institute of
Environmental Health Sciences

Purpose of the Courses

- To introduce statistics and bioinformatics concepts and terms
- To provide guidance about when a particular method should or should not be used
- We won't turn you into biostatisticians and/or bioinformaticists, however ....

Statistical thinking will one day be as necessary for effective citizenship as the ability to read and write
--H. G. Wells

Outline

- Today:
  - Study design
  - Levels of measurement
  - Numerical and graphical summaries
  - Sample size determination
- Tomorrow:
  - Estimation
  - Confidence intervals
  - Principles of hypothesis testing

Study Design Depends on the Goal

- Comparison of groups
- Comparison with control group
- Trend across dose levels and/or time
- Pre-treatment/Post-treatment change
- Estimation of incidence and/or prevalence
- Associations among several factors
- Prediction of a response from several factors
- Etc., ...

A good study design will..

- Minimize bias
- Minimize confounding
- Maximize statistical power
- Make efficient use of animals, time, other resources

Bias

- The systematic tendency to over-estimate (or under-estimate) a quantity or effect

- Example: A balance has not been calibrated recently and its readings tend to be lower than the actual mass should be.

- Example: An interim sacrifice is planned in an experiment. Only the 'sickest' animals are selected for sacrifice at that time.

Confounding Factor

- An extraneous factor related to both the treatment and the effect that may obscure or exaggerate the true relationship between the treatment and the effect.

- Classical example: There is a significant positive correlation between the crime rate and ice cream sales. Does ice cream consumption cause crime?? Does crime cause people to eat ice cream??

  What is the likely confounder here?

Power

- The probability to detect a difference or effect when it is present.

- We would like the power to be high (>80% or >90%) – more later……

Design Toolbox

- Randomization
- Blocking/Stratification
- Matching
- Control group
- Factorial layout

Design Tools: Randomization

- Each experimental unit has a chance of being assigned to any treatment.
- Example: Use random number generator to assign animals to groups rather than "haphazardly" reaching in a cage and grabbing the (slowest) animal.
- Reduces bias by distributing pre-existing differences across treatments.
- Without the involvement of a probability process, assumptions of most statistical tests are violated.

Design Tools: Blocking/Stratification

- Experimental units are grouped according to some factor(s). Treatments are assigned at random within each homogeneous group.
- Example: Group animals by body weight before randomization.
- Increases power by separating variation due to an extraneous factor from variation due to the treatment.
- The blocking factor should be related to the response, otherwise you could lose power.

Design Tools: Matching

- A particular form of blocking in which experimental units having common characteristics are paired or matched
- Examples: Twin studies, Pre-post designs
- Just like blocking, matching increases power.

Design Tools: Control Group

- A control group provides a basis for comparison. Usually, it consists of experimental units that do not receive the agent of interest, but they are treated exactly like the other groups in all respects.

- Examples: Untreated controls, Vehicle controls

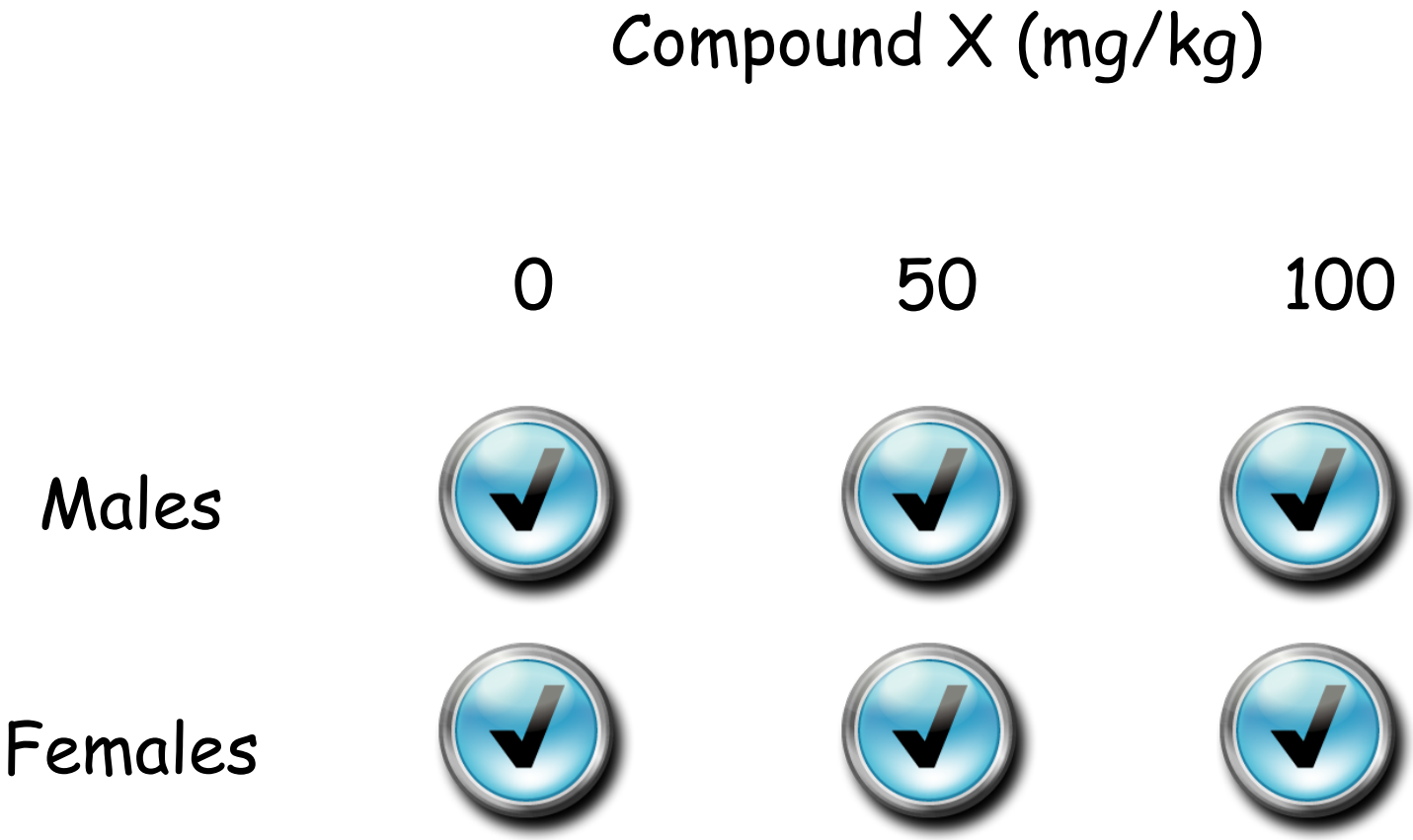
Design Tools: Factorial Layout

- The effects of two or more factors are studied simultaneously, such that all possible combinations are present in the study.

- Example: Measure the concentration of Compound X in the blood of male and female mice 1 hour after treatment with 0, 50 or 100 mg/kg of X.

- Makes efficient use of resources by allowing study of more than one factor simultaneously.

Design Tools: Factorial Layout



Design Tools: Factorial Layout

- Complete factorial: All possible combinations are present.
- Balanced factorial: The sample size is the same* in all cells.
- Incomplete factorial: Some combinations are not included. This may be done on purpose to conserve resources.

* There are exceptions to this definition.

Design Tools: Factorial Layout

- Completely randomized design
  - Experimental units are assigned at random to any one of the cells in the factorial layout.
- Randomized blocks design
  - One factor is a blocking factor (e.g., age group, gender, strain) and experimental units within each level (block) are randomized to treatments.

Further Design Considerations

- What groups/conditions/treatments are to be included?
- When/how will measurements be taken?
- How many experimental units are needed in each group?

Switching gears.....

- We've collected data, now what??
- For any data analysis, the appropriate statistical method will depend on:
  - The design of the study
  - The research question
  - What kind of data are collected

Levels of Measurement

- Nominal
- Ordinal
- Interval/Ratio

- Discrete
- Continuous

- Quantitative
- Qualitative

Levels of Measurement

- Discrete
  - Nominal — yes/no, clinical signs
  - Ordinal — grade 1, 2, 3, 4 or -, +, ++, +++
  - Count — # animals with a liver adenoma
- Continuous
  - Interval/ Ratio — mouse body weight, gene expression level

Numerical and Graphical Summaries

- Three important pieces of information:
  - Center/typical value
  - Variability
  - Shape of distribution

Measures of Central Tendency

- Mean – arithmetic average, $\bar{x}$
- Median – half-way point
- Mode – most common value(s)

- Geometric mean – log-based average

$$\bar{x}_{geom} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \exp\left(\frac{\sum \ln(x_i)}{n}\right)$$

Measures of Variability

- Variance
- Standard deviation (SD)
- Standard error of the mean (SE or SEM)
- Coefficient of variation (CV)
- Range
- Interquartile range (IQR)
- Modal percentage

- Geometric standard deviation??

Tricky! Seek professional help.

Measures of Variability

- Variance
- Standard deviation (SD)
- Standard error of the mean (SE or SEM)
- Coefficient of variation (CV)

$$Var = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}$$

$$SD = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$$

$$SEM = \frac{SD}{\sqrt{n}}$$

$$CV = \frac{SD}{\bar{x}}$$

Measures of Variability

- Range = Max – Min
- Interquartile range (IQR) = 75^{th} %tile – 25^{th} %tile
- Modal percentage = % in modal category

Standard Deviation or Standard Error?

- SD measures variability of individuals
- SE measures variability of the estimated mean

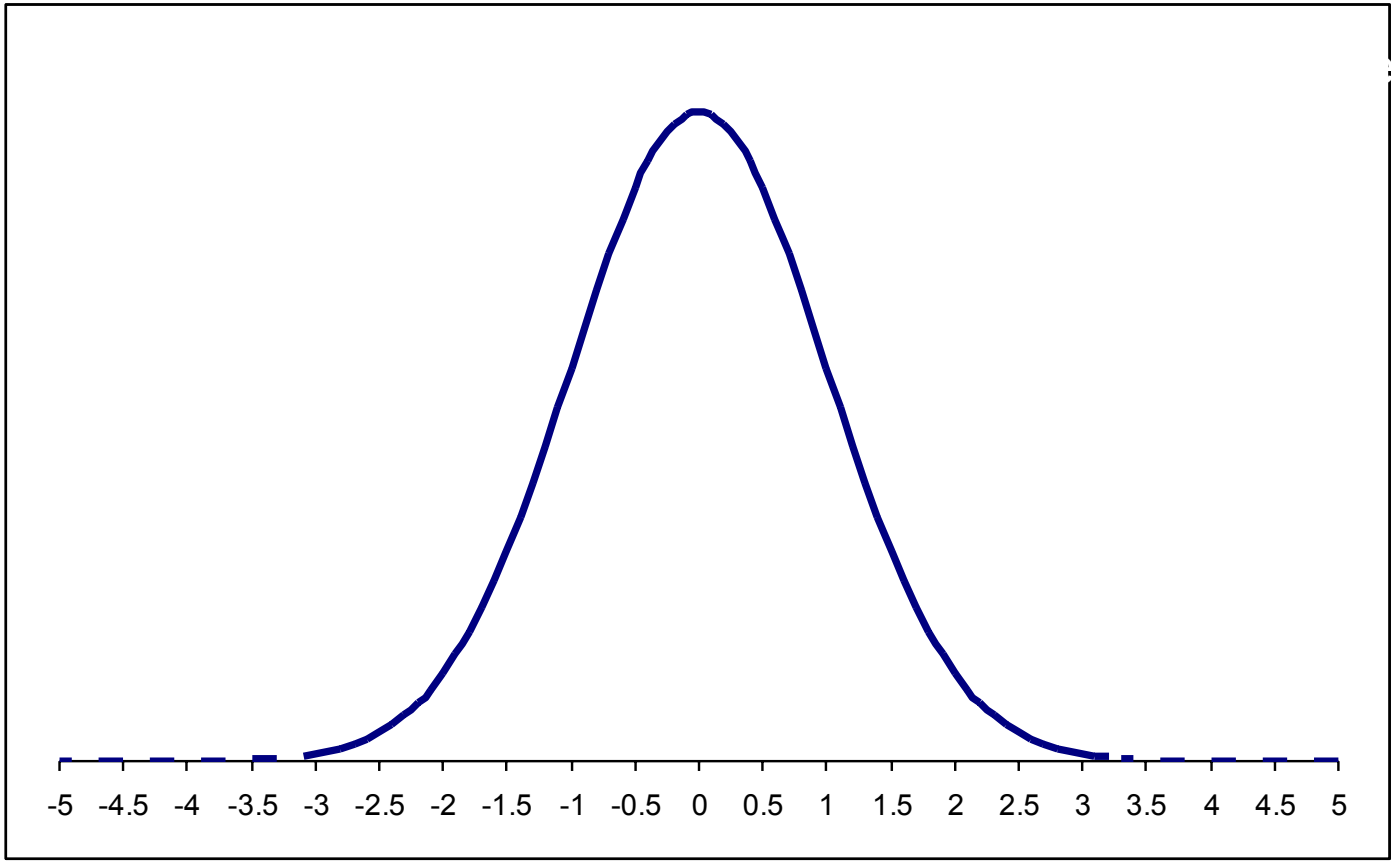
Description Depends on Measurement Level

Level	Center	Variability
Nominal	Mode	Modal percentage
Ordinal	Mode	Modal percentage
	Median	Range, IQR
Interval/	Mode	Modal percentage
Ratio	Median	Range, IQR
	Mean	SD, SE, CV
	Geometric mean	

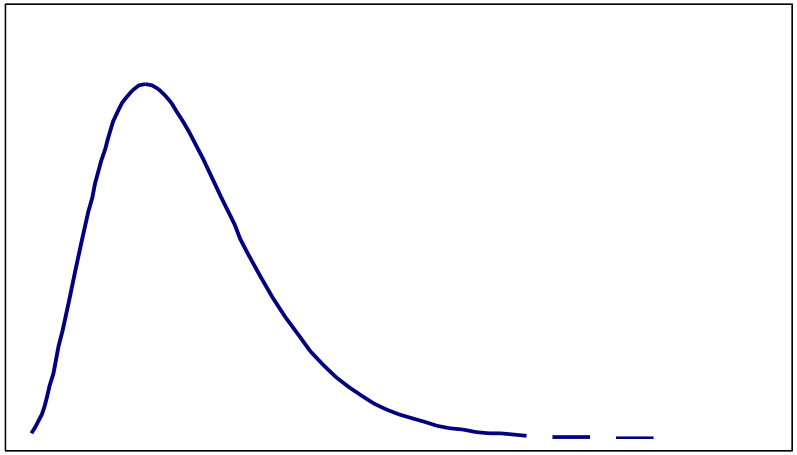
Shape of the Distribution

- Symmetric/Non-symmetric
- Skewed to right/left
- Unimodal/Bimodal/Multimodal
- Normal/Non-normal shape

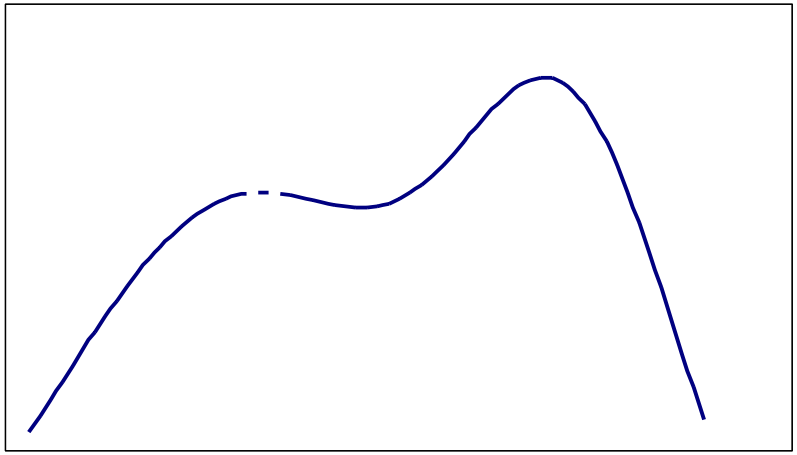
The Normal/Gaussian/Bell curve Distribution



Non-normal Distributions



Skewed to the right



Bimodal

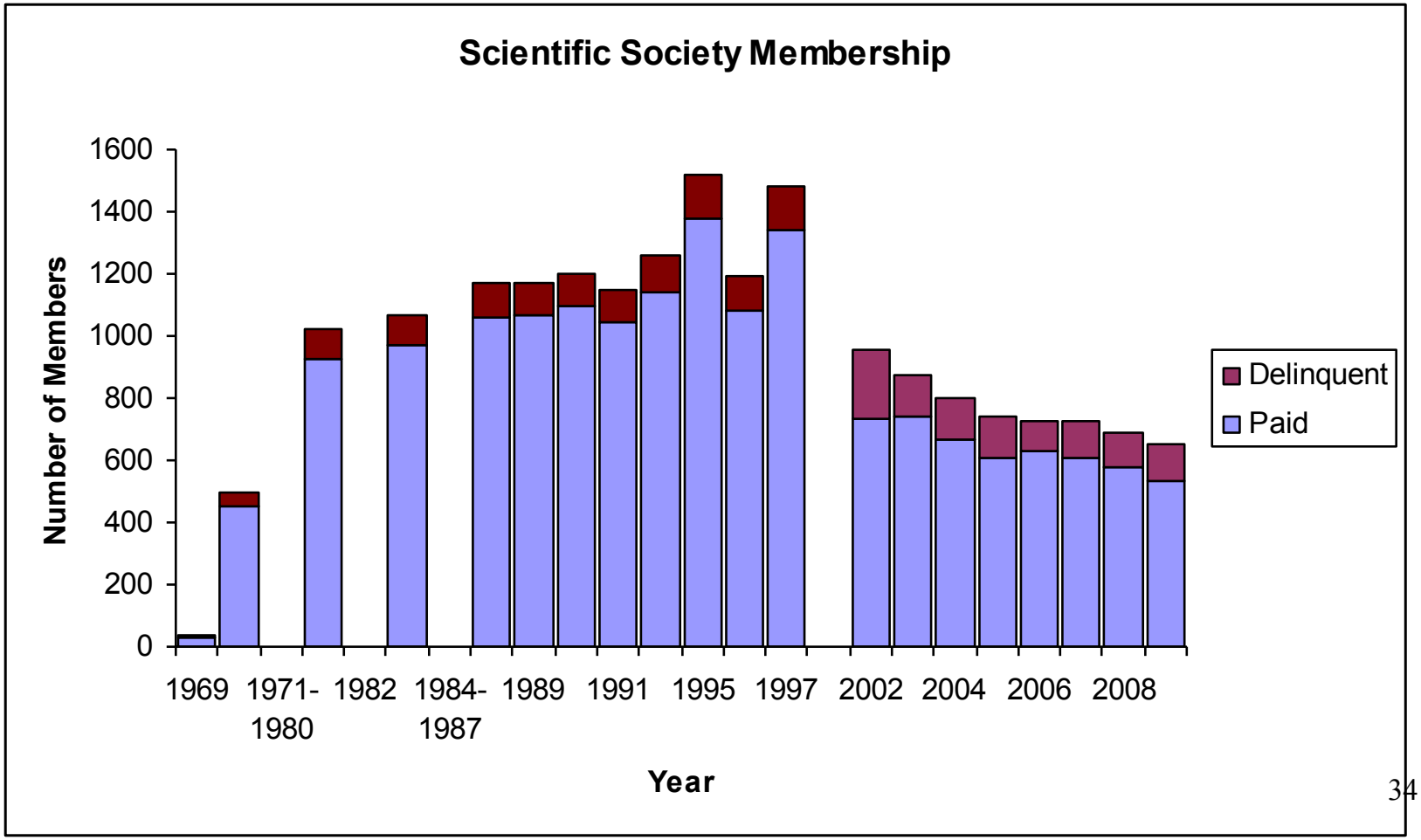
Graphical Methods

A picture is worth a thousand words!

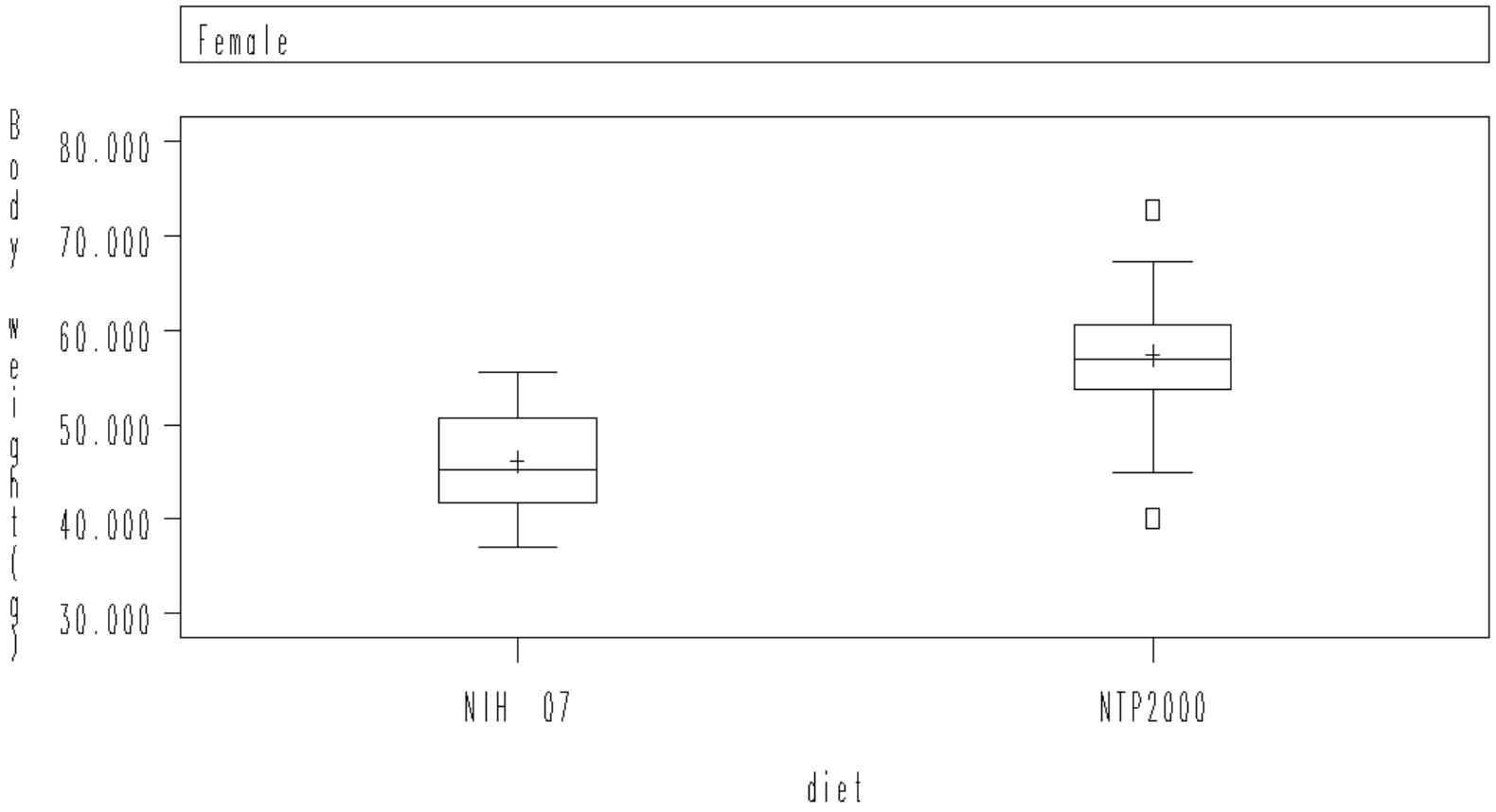
Graphical Methods

- Bar charts
- Histograms
- Boxplots
- Scatter plots
- Q-Q plots

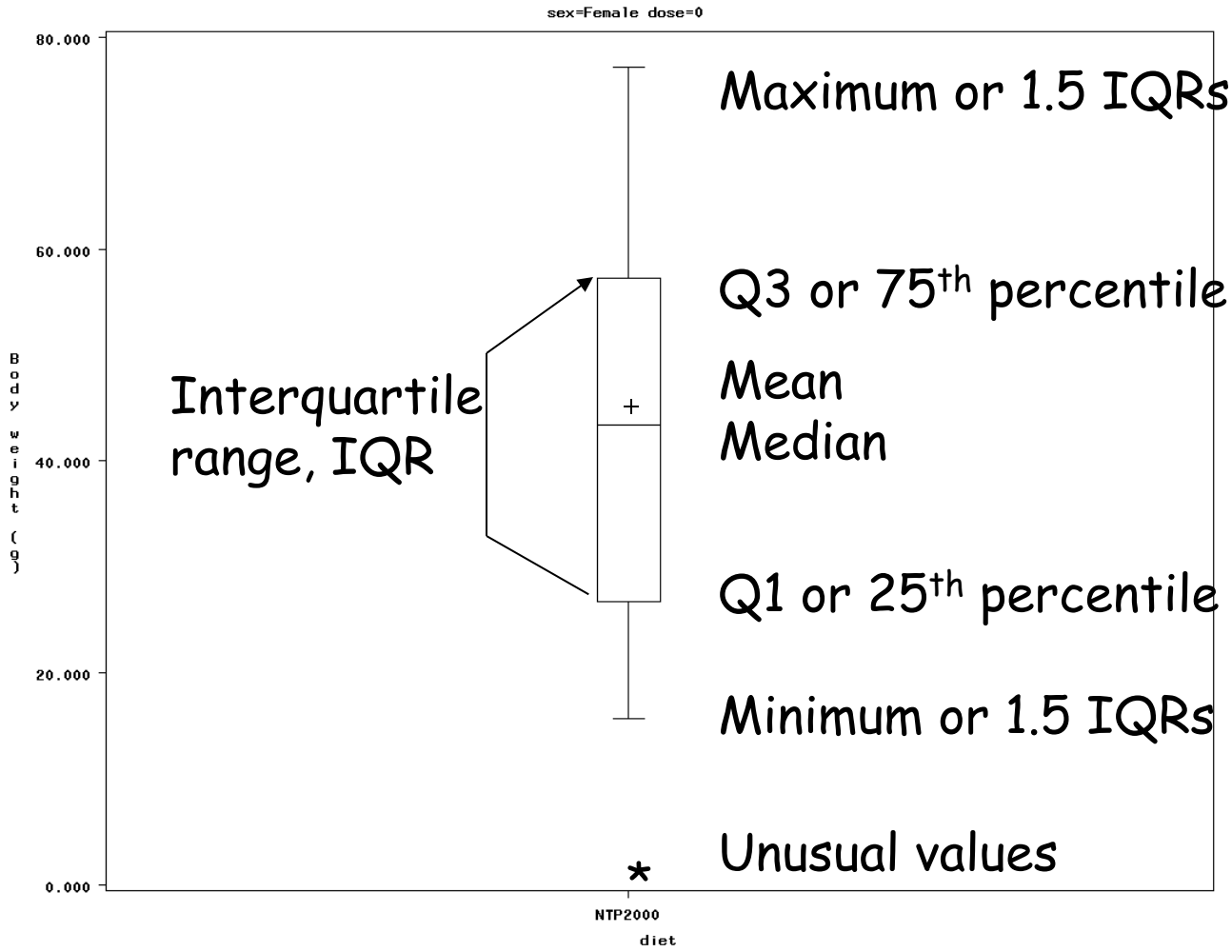
Graphical Methods: Bar Charts



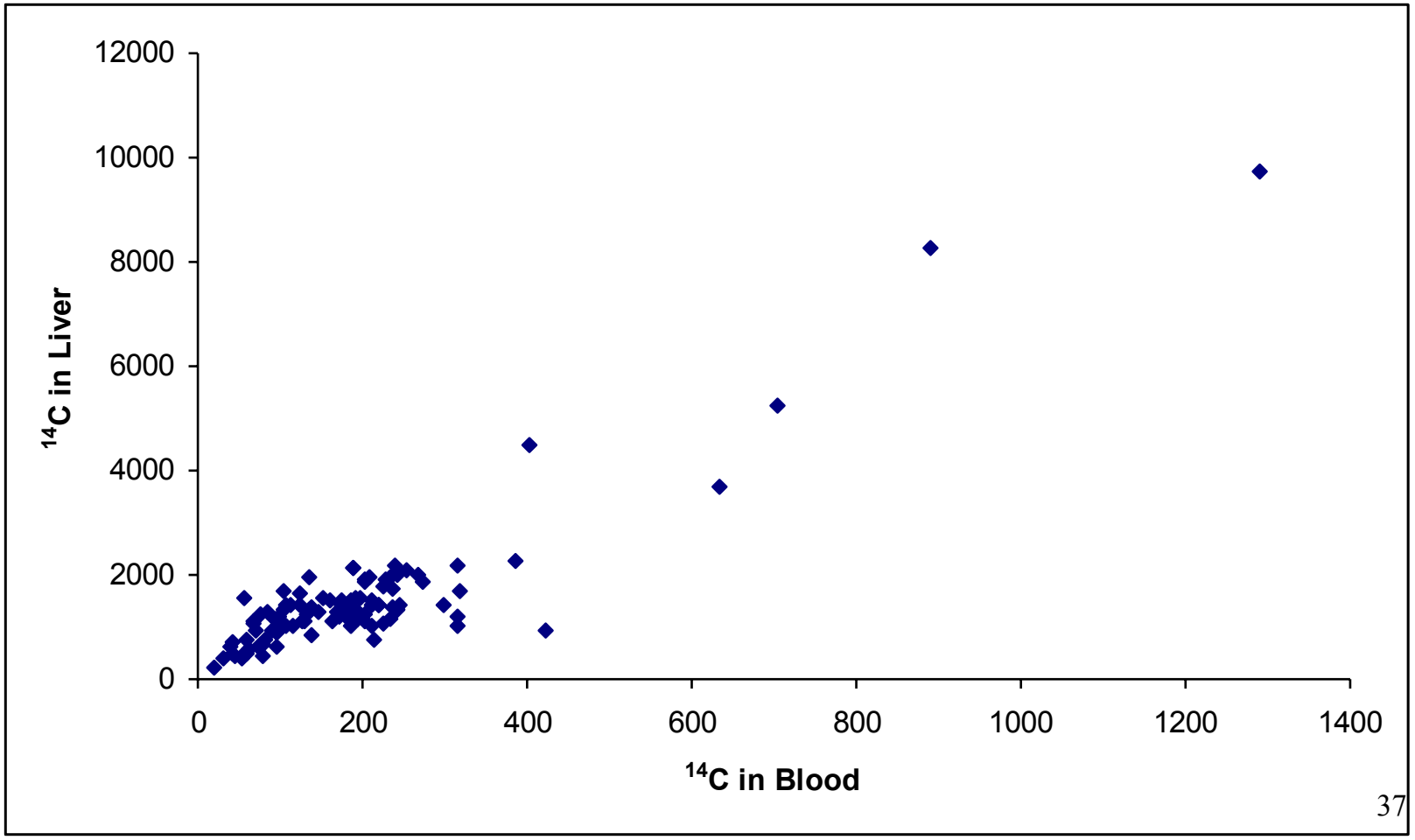
Graphical Methods: Box Plots



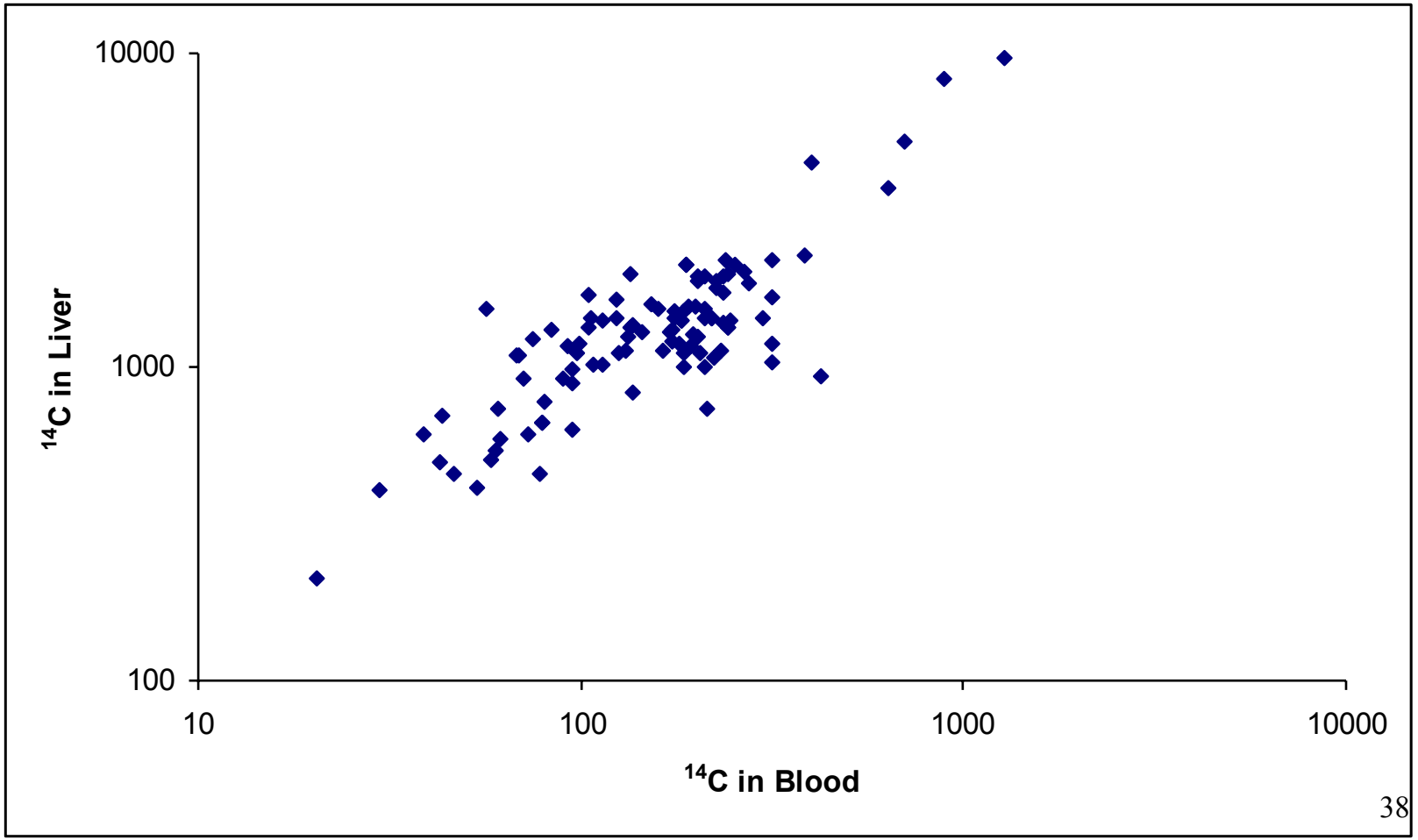
Graphical Methods: Box Plot Components



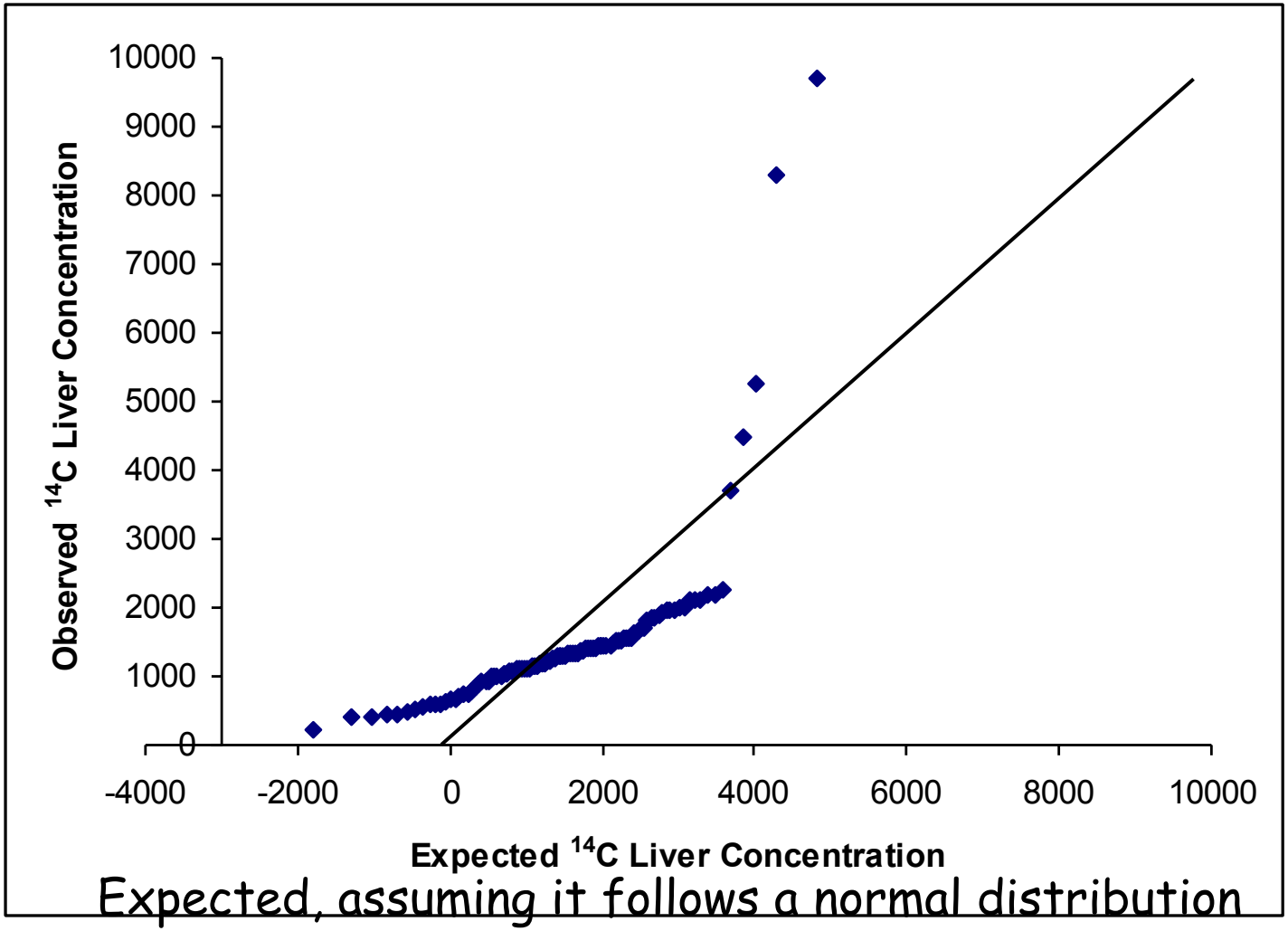
Graphical Methods: Scatter Plots



Graphical Methods: Scatter Plots

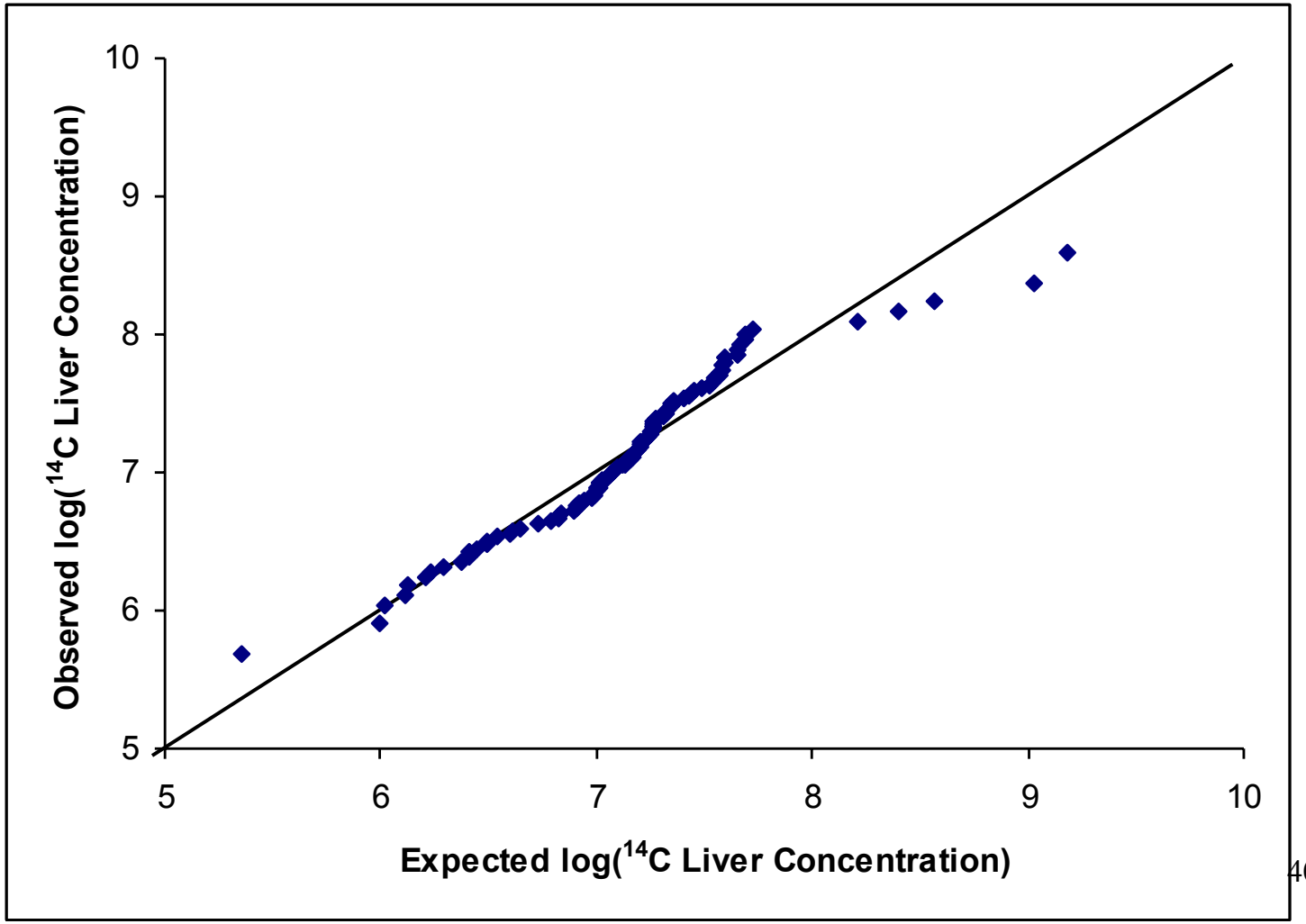


Graphical Methods: Q-Q plots (Quantile-Quantile plots)

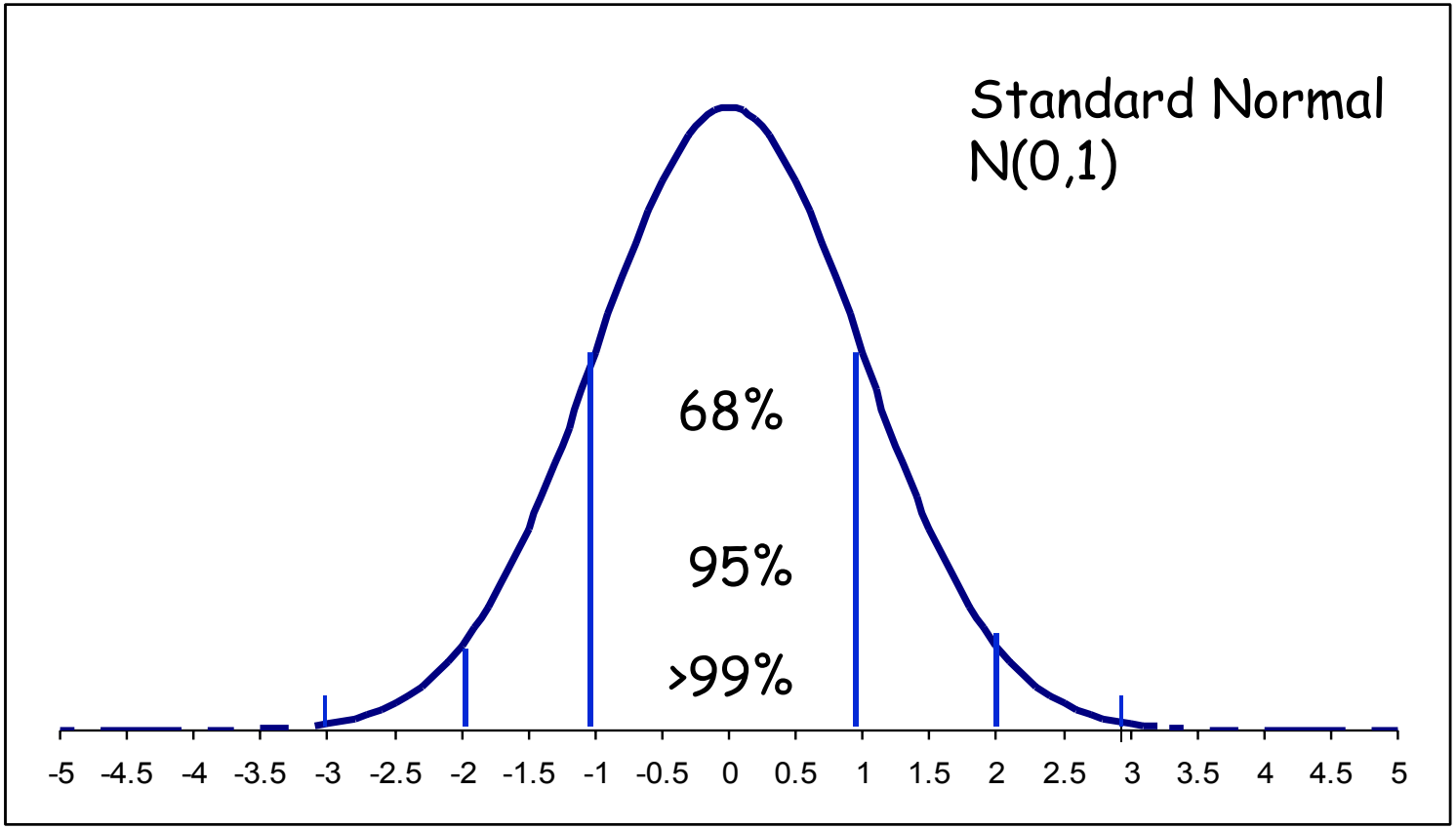


Expected, assuming it follows a normal distribution

Graphical Methods: Q-Q plots (Quantile-Quantile plots)



The Normal/Gaussian/Bell curve Distribution



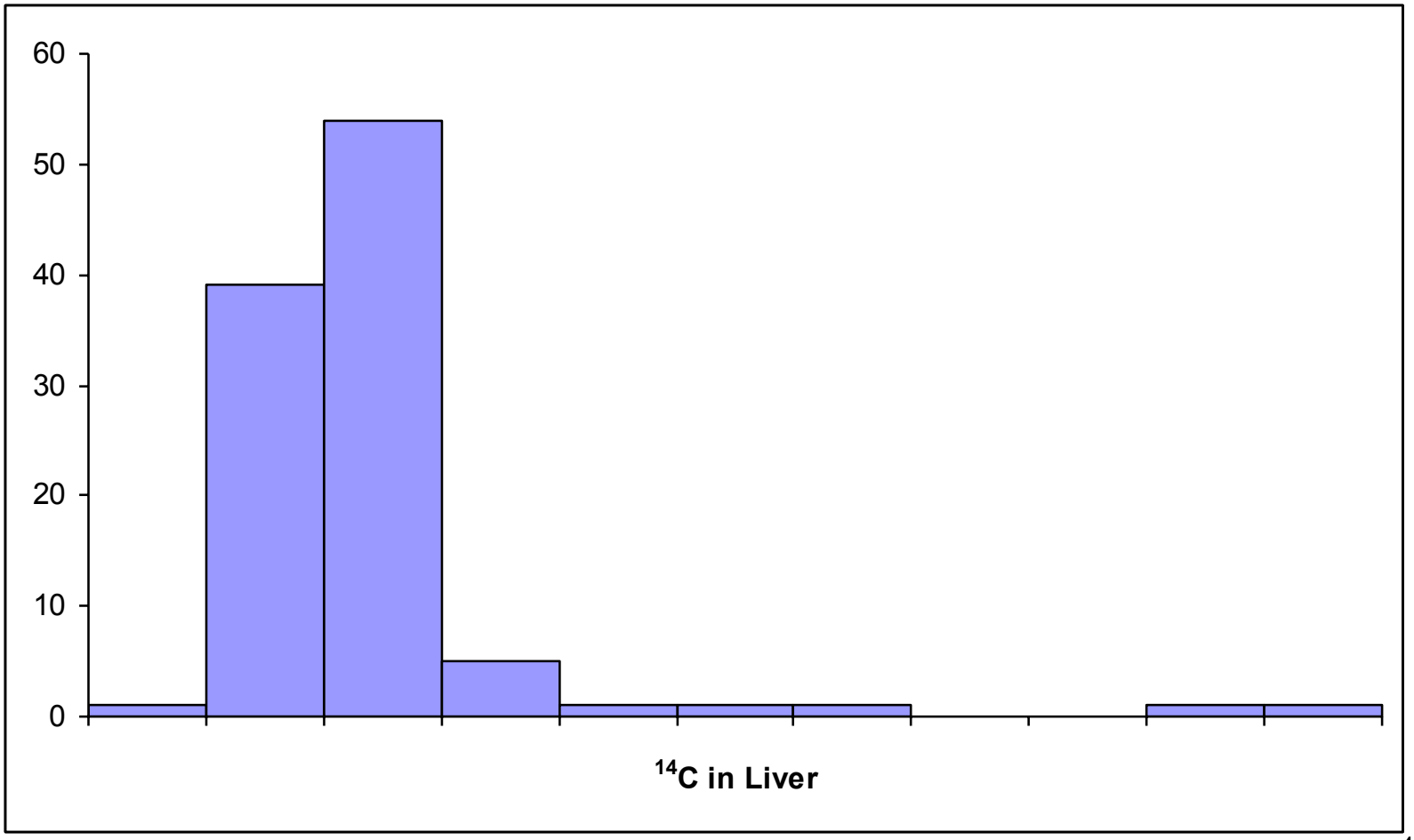
Normality and Transformations

- Many statistical methods rely on having normally distributed data (and sometimes equal variances)

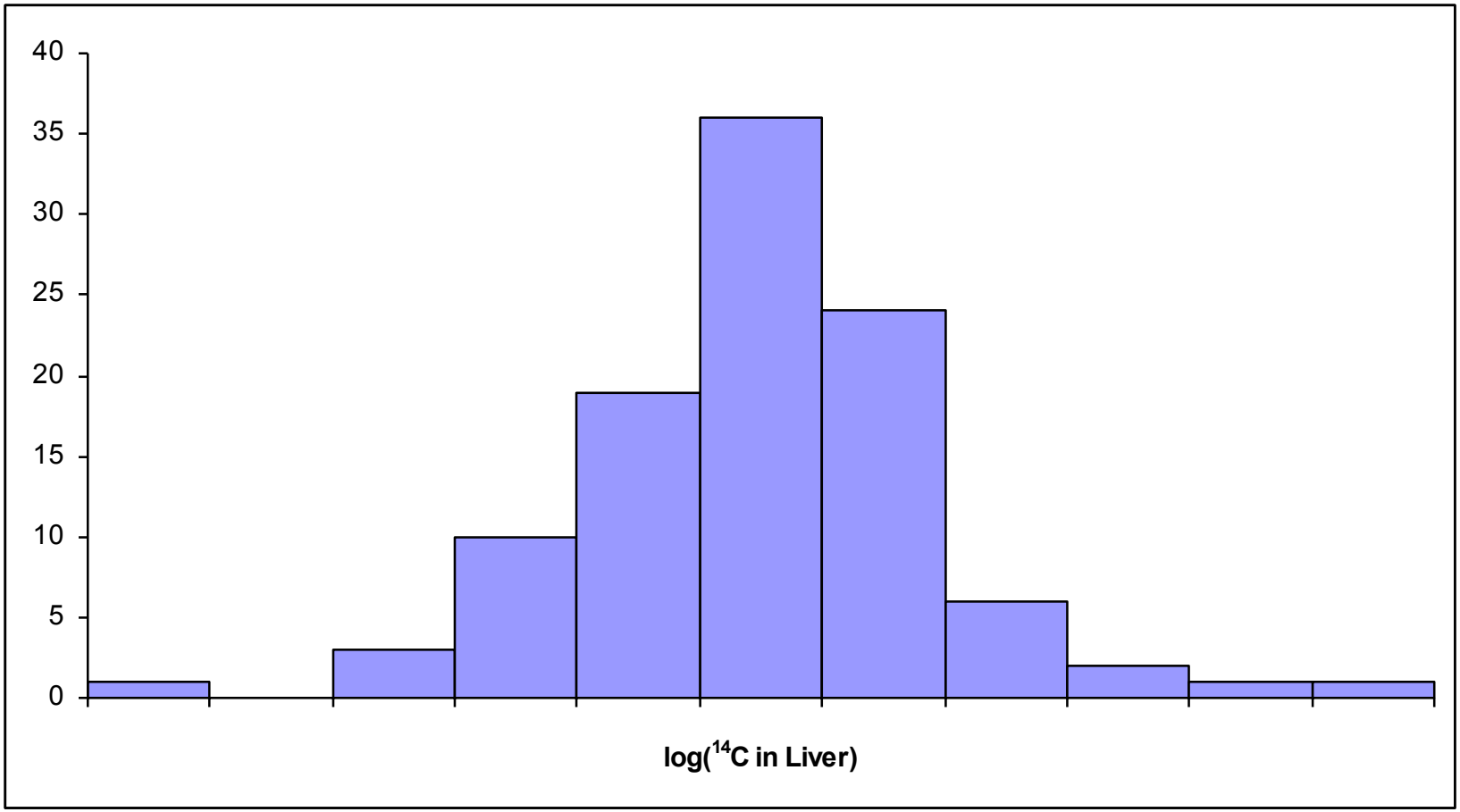
- Normalizing (and variance stabilizing) transformations

  - Logarithmic
  - Square root
  - Box-Cox power transformations, $(x^{\lambda} - 1)/\lambda$

Normality and Transformations: Liver ^{14}C



Normality and Transformations: Log(Liver ^{14}C)



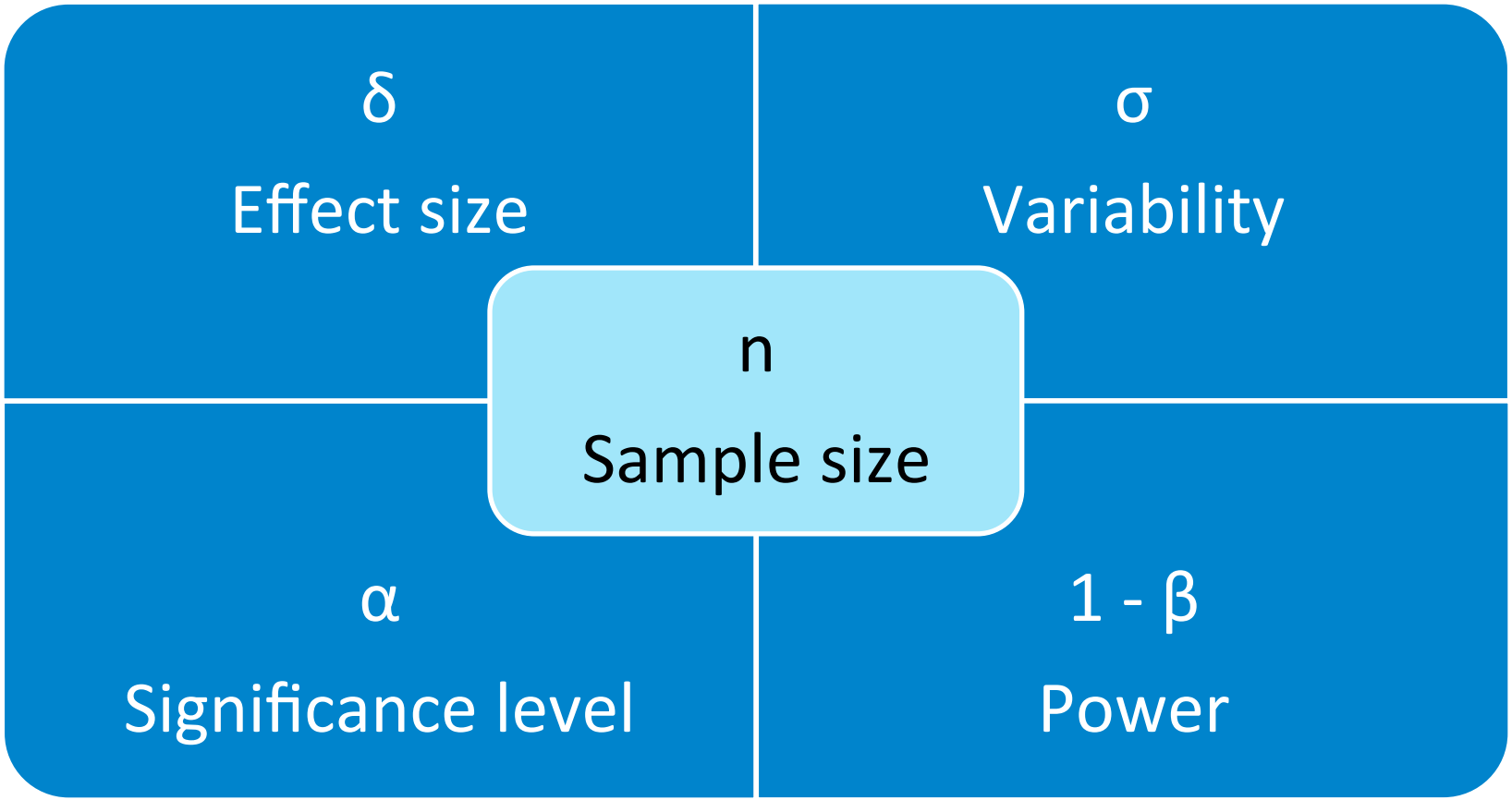
Sample Size Determination

- Depends on several things:
  - Experimental design
  - Effect size
  - Variability
  - Statistical power
  - Significance level

- Formulas, Tables, Charts, Software, Websites are available

Statpages.org is particularly comprehensive

Notation



The formula linking these together is determined by the research question and the study design.

δ, Effect Size

- Effect size is the smallest difference that you would accept as biologically meaningful or important.
  - Drug A decreases mean SBP by 0.5 mm Hg. Is that a biologically meaningful change? Probably not.
  - Drug B decreases mean SBP by 10 mm Hg. Is that a biologically meaningful change? Probably so.
  - Drug C decreases mean SBP by 5 mm Hg. Is that a biologically meaningful change?

SBP = Systolic Blood Pressure

σ, Variability

- Variability describes how much variation you expect within each group.
  - Usually estimated from pilot studies and/or previously published literature.
    - If nothing is known, σ may be estimated from the normal distribution property that the range = 6σ, so σ = range/6.
  - May be different in different groups.

α, Significance Level

- α is the probability of a false positive result.
- We typically select $\alpha = 0.05$.
- Can we predict the direction of the effect, δ?
  - If yes, we will do a one-sided test.
  - If no, we will do a two-sided test.

1 – β, Power

- β is the probability of a false negative result.
- β should be small; 1 – β should be large.
  - Typically, we choose 1 – β = 0.80 or 0.90.

Example

- Planning a study of the effects of a gene on body weight of 120 day old mice on a high fat diet.
- Two groups: wild type (WT) and knockout (KO).
- Pilot study indicates that the standard deviation of body weight at age 120 days is 3.0 g in both WTs and KOs.
- Theoretically, knocking out the gene should reduce weight gain on a high fat diet.
- Any body weight difference of 2.0 g or more is considered a "real" difference.

Example

Statistical Considerations:

- Compare 2 independent groups: WT and KO.
- Body weights are typically normally distributed.
  - We will use a t-test to compare groups.
  - Pilot data suggests equal variances in WT and KO groups.
  - The t-test will be 1-sided. Why?
- We decide on $\alpha = 0.05$ and $1 - \beta = 0.90$.

Example

Based on the research question, study design, and characteristics of the data we expect to collect, we will use this formula:

$$n = \frac{\left(z_\alpha + z_\beta\right)^2 \sigma^2}{\delta^2}$$

If we had a different research question or a different study design or non-normally distributed data or unequal variances, we would use a different formula. This is not a one-size fits all formula.

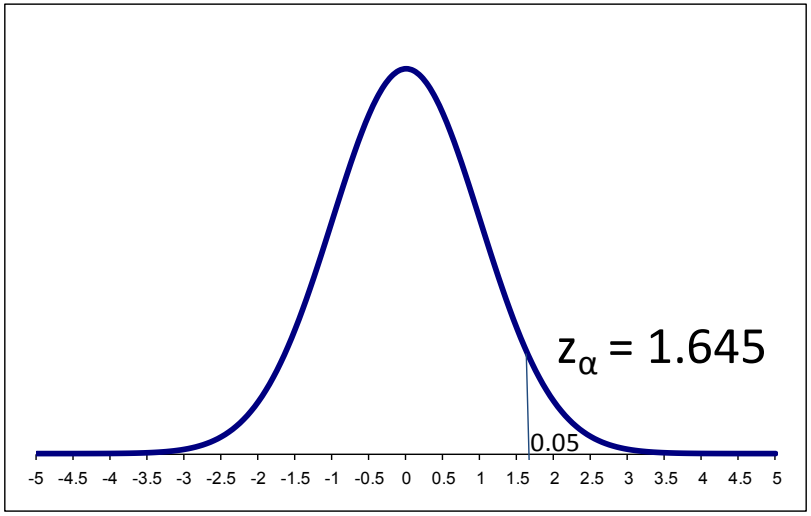
Example

$$n = \frac{\left(z_\alpha + z_\beta\right)^2 \sigma^2}{\delta^2}$$

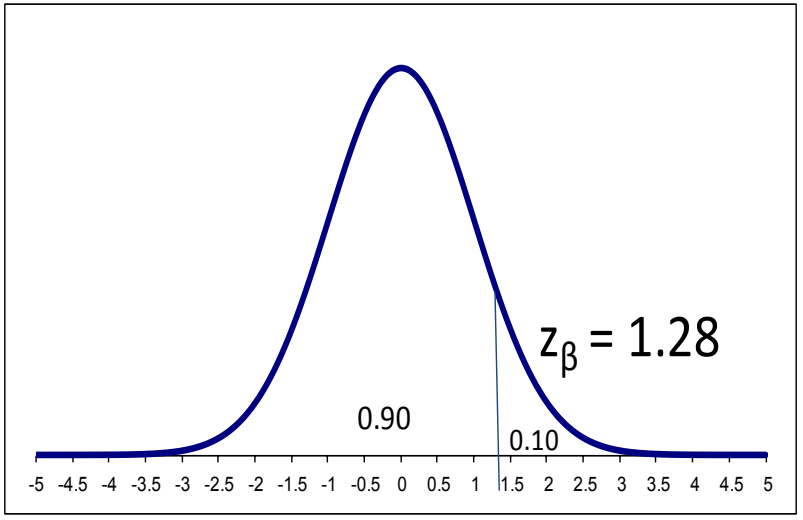
σ = 3.0

δ = 2.0

Significance Level = 0.05, one-sided



Power = 90%



Example

$$n = \frac{\left(z_\alpha + z_\beta\right)^2 \sigma^2}{\delta^2}$$

σ = 3.0

δ = 2.0

$$n = \frac{\left(1.645 + 1.28\right)^2 \times 3.0^2}{2.0^2}$$

z_α = 1.645

z_β = 1.28

$$= 19.25$$

Use n = 20 animals per group.

What if …

We cannot predict whether the knocked out gene would increase or decrease weight gain? Do a 2-sided test.

$\sigma = 3.0$

$\delta = 2.0$

$$n = \frac{(1.96 + 1.28)^2 \times 3.0^2}{2.0^2}$$

$$= 23.6$$

$z_\alpha = \sout{1.645}\ 1.96$

$z_\beta = 1.28$

Use n = 24 animals per group.

What if …

We can predict lower weight in the KOs but we're willing to have 80% power instead of 90%.

$\sigma = 3.0$

$\delta = 2.0$

$$n = \frac{(1.645 + 0.84)^2 \times 3.0^2}{2.0^2}$$

$$= 13.9$$

$z_\alpha = 1.645$

$z_\beta = \sout{1.28}\ 0.84$

Use n = 14 animals per group.

What if …

We want to compare WT and KO expression of a gene in the same pathway as the knocked out gene?

Statistical Considerations:

- Gene expression is typically lognormally distributed.
- Variance usually increases with the mean.
- We're interested in fold-changes (i.e., ratios) rather than differences between means.
- When comparing many genes, we may be more interested in the false discovery rate (FDR) than in significance and power. This is discussed in later short course classes.

What if …

We want to compare WT and KO expression of a gene in the same pathway as the knocked out gene?

In the case of expression of a single gene:

- Effect size is the fold-change that is considered biologically important.
- Variability is measured by the coefficient of variation (cv).
- The formula involves a non-central F-distribution, very messy! I'll use software for the calculations.

What if …

We want to compare WT and KO expression of a gene in the same pathway as the knocked out gene?

Suppose

δ = 2-fold difference,

σ = cv = 0.3,

α = 0.05,

$1 - \beta$ = 0.80.

Using SAS PROC POWER, n = 4 per genotype if I can predict direction of the change (one-sided test);

n = 5 if I can not predict direction (two-sided test).

Outline

- Today:

  The END.

- Tomorrow:
  - Estimation
  - Confidence intervals
  - Principles of hypothesis testing